QUESTION ANSWERING AUGMENTATION SYSTEM: CONDITIONAL SYNONYM REPLACEMENT

Kaili Wang, Yara Sevilla and Alexis Mack
Stanford University: CS224N Natural Language Processing

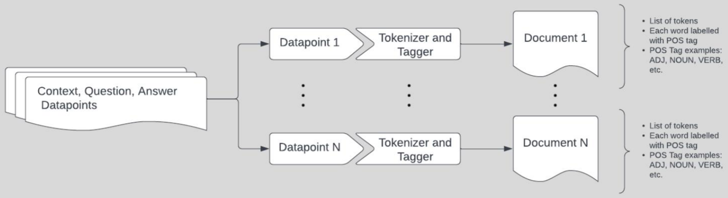
Abstract

In this project, our team implemented *Data Augmentation* tools, such that we can distinguish and target different parts of speech and bolster the data with synonym replacements. We also explore whether augmenting just the context or augmenting the context, question, and answer provides better EM and F1 results. We hope to speed up data augmentation and experiment with the number of epochs for testing and the targeted parts of speech to synonymize.

Introduction

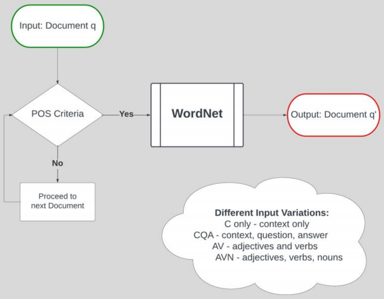
For human-computer interaction, natural language is the best information mechanism for humans. Natural language systems present issues in answering specific questions of users. The limitation inspired Question Answering Systems. Question Answer Systems aim to automatically understand and answer a question originally in a natural language context (Aroussi et al., 2016). Question Answer systems (QAS) have an advantage over search engines in meeting a user's information needs. Still the quality of Question Answering (QA) is still not sufficient regarding the number of questions that are answered correctly. The research problem is how the Question answering (QA) quality can be improved.

Data


Context, Question, Answer Datapoints
Datapoint 1
Tokenizer and Tagger
Document 1
List of tokens
Each word labelled with POS tag
POS Tag examples: ADJ, NOUN, VERB, etc.
Datapoint N
Tokenizer and Tagger
Document N
List of tokens
Each word labelled with POS tag
POS Tag examples: ADJ, NOUN, VERB, etc.

Approach

Given an input q = (context, question) we create and add an input q' = (context', question') to the data set. Here the data set is augmented with additional data rather than data replacement. When training neural networks for classification we rely on synonym replacement as a method of data augmentation. Rather than replacing generic words with synonyms we focus on replacing specified parts of speech with their synonym. We rely on the spaCY library to implement POS tagging. We rely on WordNet NLTK to retrieve synonyms.



Different Input Variations:
C only - context only
CQA - context, question, answer
AV - adjectives and verbs
AVN - adjectives, verbs, nouns

Results

Augment Strategy	EM Score	F1 Score
Baseline	47.51	30.63
Augmented Dataset A (C only, AV)	49.88	34.55
Augmented Dataset B (CQA, AV)	48.77	32.00
Augmented Dataset C (CQA, AVN)	48.77	32.00

Key Conclusions

- The EM and F1 scores for Dataset A yielded the best results. We hypothesize that this is a result of stronger correlations between synonym pairs.
- The EM and F1 scores do not vary between the two Datasets B and C. We hypothesize that this is because of indistinguishable augmentation changes.

Possible Future Improvements

- Increasing the number of synonyms given by WordNet
- Applying different weights to new data, depending on how many different words there are to the original dataset
- Exploring how these new Supervised Learning examples could be applied to Test Time Training algorithms
- Utilizing different POS tagging methods to distinguish gerunds or proper adjectives

Acknowledgements

[1]Edward Ma. NLP Augmentation, 2019.
[2]S. A. Aroussi, N. E. Habib and O. E. Beqqali, "Improving question answering systems by using the explicit semantic analysis method" 2016.
[3] Murata, Masaki. "Automatic Selection and Analysis of Verb and Adjective synonyms from Japanese Sentences using Machine Learning". 2019.